Understanding Moderators of Peer Influence for Engineering Viral Marketing Seeding Simulations and Strategies

Research-in-Progress

Daning Hu
University of Zurich
Zurich, Switzerland
hdaning@ifi.uzh.ch

Jiaqi Yan
School of Information Management,
Nanjing University
Nanjing, China
jackiejqyan@gmail.com

Algesheimer Rene
University of Zurich
Zurich, Switzerland
rene.algeshaimer@business.uzh.ch

Meierer, Markus
University of Zurich
Zurich, Switzerland
markus.meirer@business.uzh.ch

Abstract

Seeding as an emerging viral marketing strategy requires a better understanding on how various contextual factors that embedded in social networks affect peer influence and product diffusion. Realistic simulations for seeding need to incorporate empirical insights about the complexities (various moderators) and dynamics (temporal changes) of peer influence by analyzing real-world data. We analyze the impacts of peer influence moderators in a large-scale phone call network of 0.48 million customers with 364 million calls and 3.9 million video-on-demand purchases, to design empirical models and engineer data-driven simulations of product diffusion, as well as developing and evaluating seeding strategies. We intend to contribute to existing research by 1) enriching the theoretical and empirical understanding of peer influence moderators for stakeholders, 2) combining econometric models and analyses with data-driven simulations towards a complex system approach for devising and evaluating effective seeding strategies in different scenarios.

Keywords: moderators, peer influence, viral marketing, seeding

Introduction

Nowadays firms are increasingly interested in using viral marketing campaigns to promote the diffusion of their products in customer social networks. These campaigns often use seeding strategies to identify a set of individuals in social networks who can efficiently influence their friends' product adoption decisions, and thereby further trigger large scale network cascades of such adoptions (Aral et al. 2013; Aral et al. 2011; Domingos et al. 2001; Fang et al. 2013; Oliver Hinz et al. 2011). A set of conflicting seeding strategies have been proposed by different researchers, mainly focusing on network structural based measures, such as hubs (i.e., nodes with high degrees), fringes (low degrees), or bridges (high betweenness). However, how such seeding strategies may perform in different situations remains largely controversial in literature due to the complexity of various factors that may affect social contagion and the high costs associated with real world viral marketing experiments or campaigns.

Simulation as a cost-effective approach has been widely adopted to study the effectiveness of different seeding strategies in viral marketing (Abrahamson et al. 1997; Alessio et al. 2012; Aral et al. 2013; Arastoo et al. 2016; Cui et al. 2014; Damon Centola et al. 2007; Goldenberg et al.; Jankowski et al. 2013; Kempe et al. 2003; Kitsak et al. 2010; Kvasnicka 2014; Samik et al. 2010; Van den Bulte et al. 2007b). In network and computer science, this problem is also often called the influence maximization problem (Kempe et al. 2003). The majority of such seeding simulations relied on theoretical models and largely ignored the impacts of contextual factors on peer influence (and product diffusion) in various empirical settings (data sets). We suggest that such factors in the context of viral marketing mainly include customer (Aral 2010) and product characteristics (Aral et al. 2011), customer relationships (networks), as well as purchase records. This is mainly due to lack of comprehensive empirical data sets and appropriate simulation framework that can incorporate those factors.

On the other hand, there is very few empirical studies of seeding strategies since real-world network intervention is usually very expensive and can only make bounded generalizations (Aral et al. 2013). Nevertheless, we suggested that empirical insights of product diffusion in social networks can be used to inform the design of effective seeding strategies in viral marketing. However, even those empirical viral marketing studies mainly focused on using methods such as randomized experiments or econometric analysis of observational data to identify casual peer influence (Aral et al. 2009; de Matos et al. 2014; Koch et al. 2015), but largely ignored the (moderating) impacts of various contextual factors on peer influence that may have important implications for seeding strategies.

For example, when it comes to choosing which one of recently aired popular Hollywood blockbusters (e.g., Star Wars) to watch, a young man often will conform to the choice of the majority of his peers (e.g., classmates). But if he wants to find a good French artistic movie, his girlfriend who is an expert on that may have larger influence on his choice. There are various contextual factors embedded in the above social contagion scenario that may affect (moderate) the peer influence this man's friends exert on him, including age, gender, the movie type, as well as the relationship type and strength for both individuals.

Our work aims to complement existing work by combining the empirical insights derived from the analyses of such moderators with the flexible simulation approach. We first extract information from a unique large-scale phone call network of 486,000 customers in a major European broadband cable company, with the detailed records of personal demographic information and their 364 million phone calls and 3.98 million Video-on-Demand purchases over a 2 years. We selected four major types of moderators and develop an empirical model of product diffusion that can incorporate the complexity and dynamics of those moderators, including 1) product characteristics, 2) customer characteristics, 3) customer relationships, and 4) customer-product relationship factors. Then the first research question is:

RQ1: How various contextual factors in viral marketing may moderate customers' peer influence on each other's product adoption decisions?

We then use the insights derived from the above analyses along with the empirical model to design data-driven simulations of seeding and product diffusion. Moreover, we will develop seeding strategies that can incorporate the complexity (various moderators) and the dynamics (temporal changes) of peer influence. Our data-driven simulations can then be applied to compare the effectiveness of our seeding strategies with various state-of-the-art seeding strategies. So the other two research questions of our study are:

RQ2: How to use our empirical insights and model to design data-driven seeding simulations?

RQ3: How to develop and evaluate the effectiveness of seeding strategies that can incorporate the complexity (various moderators) and the dynamics (temporal changes) of peer influence?

Our work intends to contribute to marketing and social contagion research mainly in three ways. First, while previous research focused on identifying peer influence in product diffusion, our research is among the few studies to systematically investigate the impacts of various moderators of such influence by leveraging the rich information in a unique empirical dataset, thereby enriching our understanding on how social influence work under different social and economic conditions. Second, the proposed empirical data-driven simulation allows researchers and marketers to examine the effectiveness of different seeding strategies in a more realistic but also flexible way. Third, the empirical insights from our analyses and simulations can guide us to design more effective and dynamic viral marketing seeding strategies. At last, the approach developed in this work that combine both econometric analytical methods and simulations can also be applied to study other social and economic mechanisms in complex systems.

The remainder of this paper is structured as follows. In the next section, we review the studies that are relevant to this research. We then introduce our unique empirical data set and present the overall research design. The fourth section will show our empirical study that aims to discover the impacts of various moderators of causal peer influence. At last, we will discuss our ongoing work in developing a simulation approach and our own seeding strategy that incorporates the empirical insights learned from the above analyses, and compare the effectiveness of this strategy with other ones.

Research Background

Social (peer) influence refers to the impacts from the social interactions among individuals on their behaviors (Rice et al. 1990). It is an important driver that affects individuals' adoption behaviors in social networks (Ibarra et al. 1993; Leenders 2002). Aral (2011a) conceptualized peer influence based on the utility theory as "how the behaviors of one's peers change the utility one expects to receive from engaging in a certain behavior and thus the likelihood that one will engage in that behavior." In the context of product diffusion, peer influence can be defined as how the purchase of a product by one's peers changes his perceived utility and likelihood to buy that product. In other words, peer influence is a type of "direct causal influence" where a focal behavior (e.g., adopting a product) will cause the same behavior in one's friends.

Several approaches to identifying the impacts of peer influence on individuals' behaviors were developed in recent years. Tucker (2008) investigated the technology adoption behaviors in a communication network within an investment bank using instrument variables approaches. He used exogenous shocks on the benefits of using video-message technology to identify causal effects among employee's adoption decisions. Ghose et al. (2011) used structure models to study how mobile phone users' content generation behaviors relate to their content usage behaviors. They found that increasing content usage in the previous time period will negatively affect current content generation and vice versa. However, these studies mainly focused on the identification of influence and largely overlooked the confounding effects of homophily, not mentioning quantitatively analyzing the impacts of both social mechanisms on correlated individuals' behaviors.

Possible Moderators of Peer Influence

In this study, we have suggested several key contextual factors that may moderate peer influence in the context of product diffusion based on findings from previous social influence studies. We discussed how these possible moderators may work in details as follows.

Product Characteristics Aral (2010) suggested product characteristics can enable or constrain the degree to which individuals are influenced. There is a growing stream of literature that aims to investigate the characteristics of a product that make it viral. Berger et al. (2012) found that news stories that evoke high-arousal positive or negative emotions is more viral. Heath et al. (2001) found that disgusting stories are more likely to be shared by people. It was also suggested product price may be an important candidate as moderators of peer influence (Aral 2011b). Wang et al. (2015) investigated the moderating effects of two book characteristics - age and popularity - using data from a popular online rating web site in China.

Individual (Customer) Characteristics Another stream of research found that individuals who are more influential than other and can catalyze the diffusions of products, opinions, and behaviors (Coleman et al. 1957; Katz et al. 1955; Valente 1995; Van Den Bulte et al. 2007a). The complementary mechanism is an individual's susceptibility to influence (Valente 1996; Watts et al. 2007). More recently, Aral et al. (2012) found that both mechanisms are affected by individual characteristics and determine the diffusions of behaviors. Based on the above studies, we suggested that it is important to understand how customer characteristics make one individual more influential in product diffusion.

Customer Relationship (Network) Characteristics Wellman (1997) suggested that the strength of the social ties will have impacts on the social influence transmitted through such ties. The assumption is that strong ties are based on investment of time and efforts. Thus people's decision making tend to be influenced by each other. In our study, we then conjectured that people with stronger ties have greater peer influences on each other (Levy 1992; Levy et al. 1993). More recently, Aral et al. (2014) examined the impacts of two types of relationship characteristics - network embeddedness and tie strength on peer influence. They suggested that the theoretical distinctions between different types of tie strength have not been well developed to estimate how they each may moderate peer influence. Moreover, Wejnert (2002) suggested that network structures may also affect individuals' opinions and behaviors. Wellman (1997) also pointed out that people who are connected through various relationship patterns may develop common opinions and behaviors.

Relationship Characteristics between Customers and Products Centola (2011) suggested that individuals with similar past product preferences may share similar opinions on a new product. The assumption is that the more similar two individuals' past product purchases are, the more likely they will adopt similar products in the future. However, few studies has examined the impacts of such relationship characteristic on peer influence. We suggested that characteristics like the number of products a customer bought, the similarities among customers' purchases may be peer influence moderators.

To summarize, to develop realistic simulations of product diffusion, empirical insights from systematic study on the moderators of peer influence is needed. However, existing empirical research mainly focused on the identification of peer influence. The analyses on moderators are ad-hoc due to lack of comprehensive empirical data and appropriate models that can incorporate the complexity and dynamics of peer influence.

Research Design (with Current Progress)

To address the three above research questions, we propose a research design that combines both empirical and simulation approaches including: the high dimensional propensity score matching (HDPSM) method for econometric identification of casual peer influence, an empirical study for analyzing the impacts and dynamics of moderators for such influence, as well as data-driven simulations that enable development and comparisons of effective seeding strategies under various complex realistic scenarios. Figure 1 illustrates our research design which mainly consists of four steps to 1) use HDPSM to preprocess data sample to reduce model dependence and avoid potential bias, 2) empirically model and analyze the impacts of the moderators of the changing peer influence, 2) model product diffusion based on empirical insights, and 3) develop simulations and seeding strategies accordingly.

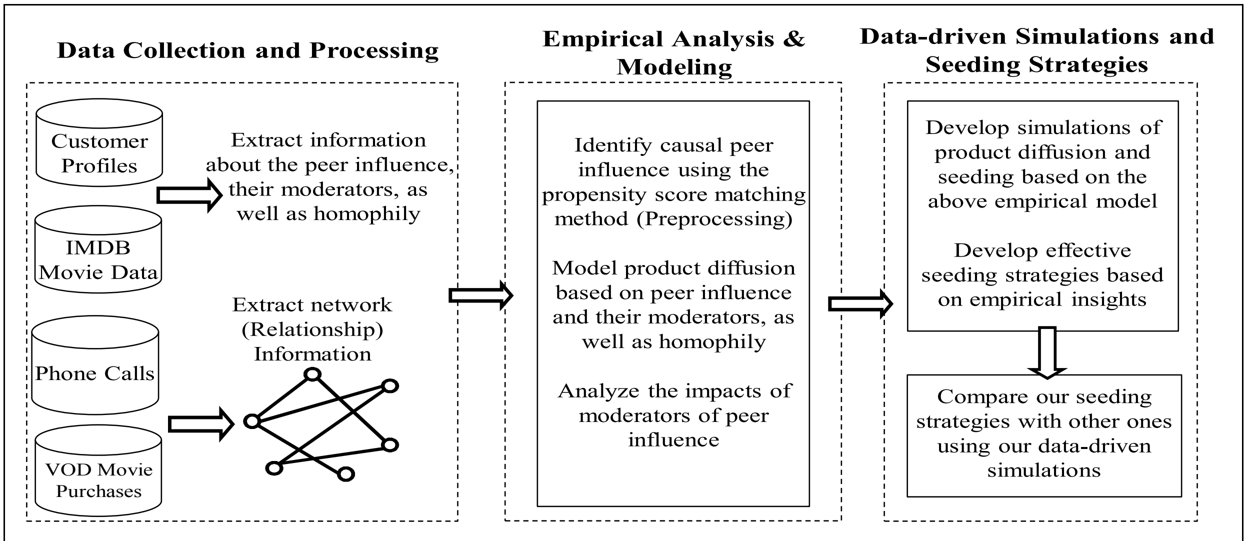

Figure 1. The Overall Research Design

There are mainly three modules in our design, indicated in the three boxes with dotted lines. The first module – data collection and processing aims to extract and prepare both the moderator information and the social network information from our unique empirical data set. The second module then uses propensity score matching method as a nonparametric preprocessing step to prepare data for parametric causal inference analysis (Ho et al. 2007). This can effectively reduce model dependence and potential bias. We then empirically analyze the impacts of peer influence and its moderators. Then a data-driven simulation model of product diffusion can be built based on the above empirical findings about peer influence and its moderators, as well as homophily factors. In the third module, we can develop simulations and seeding strategies by incorporating the impacts of peer influence and their moderators. At last, we will compare the effectiveness of these seeding strategies with several state-of-the-art ones in our simulations. In the next few sections, we introduce each of the modules in detail.

Data (Research Testbed)

The data for this study is from one of the largest broadband cable operators in Europe. It mainly offers its customers three types of services: telephone service, TV programming (including a Video-on-Demand service), and Internet. The VOD product selections include movies (most are Hollywood productions and some European movies) and TV show episodes, as well as a few concerts and adult films. The VOD purchase is charged on a pay-per-view basis and grants the customer unlimited access for 24 hours. Each customer has a unique ID number that is used for all of their activities across different types of services.

We have gained access to the company's anonymized customer data, phone records and VOD transactions, for a 2-year period from 2012 to 2013. The anonymized customer demographic information includes anonymized ID, gender, birth year, and preferred language for contact. Such information can be used to study homophily and its impacts on product diffusion. The phone usage data contains customers' outgoing phone call information, including phone numbers (coded for anonymization purposes) of the callers and receivers, call types, the length, frequency and cost of the calls, etc.

The unique phone call data can be used to construct a social network whose relationships sometimes are more personal than online social contacts (e.g., Facebook friends) which are intensively studied by existing research. The phone calls capture people's offline social interactions and often indicate they know each other offline. In daily life, it usually is more difficult to obtain a person's phone number than his online contacts. Moreover, phone calls have been used to study social relationships among individuals in many studies (Akbaş et al. 2013; Eagle et al. 2009; Zignani et al. 2015). Eagle et al. (2009) found that it is possible to accurately infer 95% of friendships based on the observed phone call data. (Akbaş et al. 2013) also analyzed phone interactions among individuals to generate their social networks. These findings have provided empirical evidence that phone calls can reflect the real world social relationships among people.

The VOD transaction data provides information about videos the customers purchased, including video title, purchase date, and cost. Moreover, we have use web crawlers to automatically collect information about the movies in the VOD data set from the Internet Movie Database (http://www.imdb.com/). Such information along with the demographics and VOD information provide good indicators for customers' movie tastes. Moreover, in compliance with local telecommunication and data privacy regulations, the data provider replaced the names of customers with unique customer IDs to anonymize the data. Table 1 shows the basic statistics of the phone call and VOD data.

Table 1. Basic Statistics of the Phone Call and VOD Data

01/01/2012 - 12/31/2013	Number of customers /nodes	Number of outgoing calls	Number of unique phone number pairs (pairs within company)	Number of purchased videos
Complete data	486,002	364,161,136	58,918,755 (1,077,546)	3,983,757
Extracted phone network data	85,592	78,479,709	15,872,931 (241,201)	2,165,162

To analyze the impacts of various contextual factors embedded in customers' social relationships on their VOD purchasing behaviors, we extract phone calls of VOD customers from the complete data set to construct a social network. Such a phone call network is different from most existing empirical viral marketing studies that mainly focused on online social networks (e.g., Facebook) since it captures people's offline social and economic (work) interactions. We conjecture that phone calls between two individuals often represent a stronger personal relationship than online interactions. This is because the anonymity nature of the Internet provides people extra protections of their privacy and identity. In daily life, it usually is more difficult to obtain a phone number than an online contact. In the phone call network, each customer node has ordered at least one video, and all nodes included are customers of the company. A pair of nodes are friends if one has called the other at least once. The constructed network consists of 85,592 customers with 241,201 relationships, involving 78.5 million calls and 2.2 million VOD purchases.

High Dimensional Propensity Score Matching (HDPSM) as Preprocessing

Previous research (Aral 2010; Aral et al. 2009; Aral et al. 2013; Aral et al. 2011) have proposed that social contagion is mainly driven by two important mechanisms peer influence and homopihly. Aral et al. (2009) proposed to use HDPSM to econometrically identify and estimate the impacts of the causal peer influence (versus homophily) on users' adoption decisions of a Yahoo mobile service application. Their results show that peer influence (versus homophily) is overestimated in the early stage of the diffusion of this Yahoo application and later being reduced. Based on such empirical findings, they further develop a simulation model which incorporate these two mechanisms. However, many of the contextual factors and their impacts on peer influence were not investigated due the limitation of their empirical data set. For instance, the heterogeneity in the product characteristics is not considered since their data only records the diffusion of a single applications instead of multiple ones.

We also adopt and extend the HDPSM analytical framework used in (Aral et al. 2009; Aral et al. 2013) to our unique data set. In our empirical study, the treatment for a customer who might purchase a specific video M is defined as having one or more friends in his phone call network who have purchased M. We then match treated customers with untreated individuals that were as likely to have adopter friends at the selected time period (i.e., the control), conditional on a vector of 33 observable characteristics and behavioral attributes (X_{it}). Due to the length limitation of the RIP paper, the details of the 33 characteristics are not reported here but included in the appendix (Table 1). Such matched individuals can be used to estimate the part of perceived peer influence effects that actually is originated from the homophily effects. For each selected time period t, we calculated p_{it}, the propensity for an individual to have been treated at time t, using a logistic regression as below: $p_{it} = P(T_{it} = 1 \mid X_{it}) = \frac{exp[\alpha_{it}+\beta_{it}X_{it}+\varepsilon_{it}]}{1+exp[\alpha_{it}+\beta_{it}X_{it}+\varepsilon_{it}]}$ (1), where T_{it} is the treatment status of customer i at time t and X_{it} is the vector of 33 individual characteristic and behavioral covariates of i. Logistic regression is used because it is the standard technique in HDPSM.

Preliminary Findings of the HDPSM Analysis

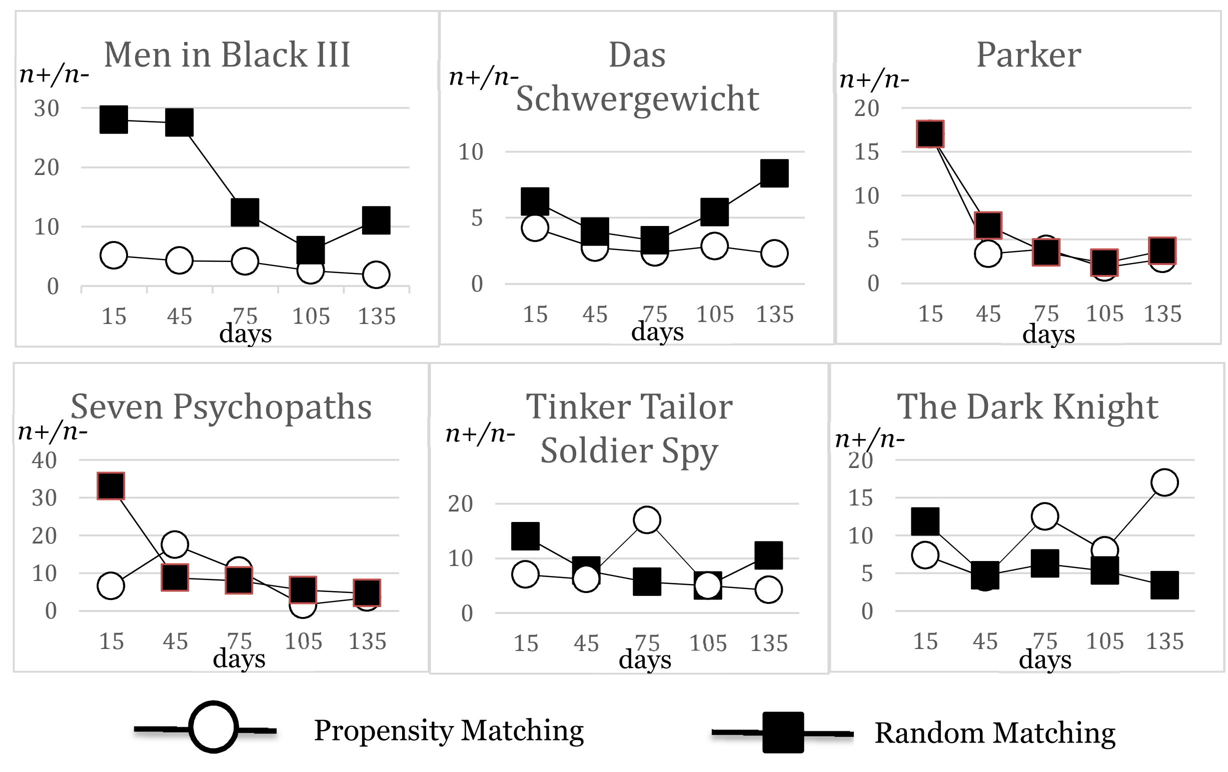

Figure 2. HDPSM Analysis Results for Selected VOD Movies

For each video at the selected time period, HDPSM generates a group of matched pairs who have very similar likelihood to have adopter friends due to observe homophily. Treatment outcome is the ratio n_+/n_- of the number of treated adopters n_+ (i.e., number of customers who have purchased the same focal video as their friends did) over the number of untreated adopters n_- in the propensity score matched sample. We then compare the treatment outcome estimated by propensity score matching (n_+/n_-) and by random matching (n_+/n_-^r), which ignores all similarity among customers, to estimate the magnitude of the homophily effects on customers' correlated video purchases.

Figure 2 shows the treatment outcomes – the fraction of observed treated to untreated adopters under propensity matching and random matching for six popular VOD movies over five time periods (from 0 to 135 days). These time periods are empirically divided to observe how peer influence and homophily change over different stages of the diffusion of a VOD movie. The results for the random matching (filled square) show the aggregated impacts of both mechanisms while propensity score matching (empty circles) indicates of the impacts of peer influence by controlling all observable homophily factors.
For instance, in the result panel of the movie "Men in Black III", during its first 15 day after release, the number of treated adopters is 5 times (filled squares) of the randomly chosen (i.e., uncontrolled homophily) untreated adopters. This means the aggregated impacts of homophily and peer influence is estimated to increase the likelihood of product diffusion by five times. In the meantime, treated adopters is 28 times more than the number of propensity matched untreated adopters. This means that, after controlling homophily, the impacts of peer influence alone is much less than homopihly during the selected time period. We can observe that the aggregated effects decreases drastically over time (from day 45 to day 75 and ahead) while the impacts of peer influence remain stable indicating the effects of homophily decreases very fast over time for the movie "Men in Black III". Our finding for this movie is consistent with the diffusion of a Yahoo application in (Aral et al. 2009), indicating homophily is the driver of diffusion in the early stages.

However, our unique data set allow us to examine the diffusions of multiple products rather than the single Yahoo application in (Aral et al. 2009). We found that the patterns for the impacts of peer influence are drastically different across different types of products. For example, the results show that the diffusions of the movies "Das Schwergewicht (Here Comes the Boom)" and "Tinker Tailor Soldier Spy" both have an increasing impacts of homophily at a late stage of their life cycles (from day 105 to 135). Moreover, at the later stage of "The Dark Knight", peer influence becomes the driving force. In addition, for the lesser known "Parker", its diffusion is largely driven by influence instead the homophily through its whole life cycle. The differences in these selected movies show that the product diffusion patterns can be drastically different due to various contextual factors embedded in the viral market scenarios.

An Empirical Model of Product Diffusion and Moderator Analysis

As Ho et al. (2007) point out that matching alone is not a method of estimation and requires an empirical model afterward to estimate the impacts of the casual and other control factors. They proposed to use propensity score matching as a preprocessing strategy followed by a parametric model. The main advantages are to reduce model dependence and adjust potentially confounding variables nonparametrically to avoid potential bias. In this study, we adopt this strategy to use HDPSM as a preprocessing step for our own parametric empirical model of peer influence and their moderators for product diffusion. Based on our empirical data set and the four major types of moderators, we develop and estimate the following parametric model based on the widely used linear-in-mean social interaction model (Brock et al. 2001; Manski 1993). For a selected (observation) time period t,

$$Watched_{ijt} = \alpha + \beta_1 AvgFrdWatched_{ijt} + \beta_2 AvgFrdWatched_{ijt} \times Mod_{kt} + \beta_3 Cus_{it} + \beta_4 Prd_{jt} + \beta_5 NET_{it} + \beta_6 CP_{ijt} + \varepsilon_{ijt} \quad (1)$$

In our empirical model, the dependent variable $Watched_{ij}$ is a binary indicator if customer i has watched VOD movie j until the end of the observation period. For each i, we examine her phone call network and identify her friends who have watched the same movie j before and thus may exerted peer influence on her. Then the measure of peer influence in our model is constructed as the average number of i's friends who have watched movie j, $AvgFrdWatched_{ij}$. Cus_{it} is the set of customer specific variables at time t, such as customer i's age, gender, or network position. Prd_{jt} is the set of VOD movie characteristics, like movie genres and sales. NET_{it} is i's network characteristics, CP_{ijt} is a set of customer-movie specific contextual variables such as $Recency_{ijt}$ which indicates days from the time i's last friend watched movie j.

Mod_{kt} is a moderator of peer influence (K is the set of moderators). If Mod_{kt} is a dyad level moderators between customer i and his friend n (N_i is the number of i's friends), equation (1) can be written as (2):

$$Watched_{ijt} = \alpha + \beta_1 AvgFrdWatched_{ijt} + \beta_2 \sum_{k\in K} \sum_{n=1}^{N_i} Watched_{njt}\, Mod_{kt} / N_i + \beta_3 Cus_{it} + \beta_4 Prd_{jt} + \beta_5 NET_{it} + \beta_6 CP_{ijt} + \varepsilon_{ijt} \quad (2)$$

Data-driven Simulations of Product Diffusion

Based on the above empirical model and the four-determinant seeding simulation model of social contagion in (Oliver Hinz et al. 2011), we develop a dynamic seeding simulation with two stages:

1) In the (start) time period t, for seed m and a movie j, the likelihood of m will watch and promote j is $Watched_{mjt} = I_{mjt} \times P_{mjt}$, where I_{mjt} is the information probability that the seeding campaign initiator that makes m aware of j, and P_{mjt} is the probability that m will participate the campaign.

2) In the following time periods, the movie j will diffuse from the seeds and other watched customers to their friends, given by equation (1).

The main advantage of our data-driven simulation approach is based on its more realistic and empirically estimated model and parameters than various pure theory-driven models. The adoption probability can be calculated based on the observed peer influence and its moderators based on empirically observed values. For instance, instead of having step-by-step network cascading models which often assumes homogeneity in customers and relationships for diffusion speed and patterns, our empirical model captures the heterogeneity in product, customer and relationship characteristics. We suggested such heterogeneity may have caused the distinct patterns of peer influence and homophily across different movies as shown in Figure 2. This allows us to more realistically simulate the product diffusions among different customers with different speeds and patterns at selected time periods.

Development and Evaluations of Effective Seeding Strategies

The next step is to develop effective seeding strategies based on the insights we learned from the above empirical analyses. But right now, such data driven empirical seeding strategies are still work in progress and not reported in this paper. Then we compare the seeding strategies developed based on the insights from our analyses and empirical model with several state-of-the-art seeding strategies. Currently, these state-of-the-art strategies serving as benchmark include seeding that targets to hubs (high-degree), bridges (high-betweenness), fringes (low-degree), or random customers. In the future, we may add more advanced link analysis based ranking methods such as HITS or Google PageRank.

In our evaluation experiments, for each movie m, the "naturally occurred" diffusion of m in our dataset serves as the baseline effects. The seeding (i.e., making selected individuals to watch m and participate the campaign to influence their friends' adoption decisions of m through incentive schemes) is the interventions added to m's "naturally occurred" diffusion.

We use two experiment settings. The first setting only allow seeding at the first time period (T_1). The second setting allows seeding at each time period. In the first setting, for each time period T_n , we use ΔT_n to represent the number of individuals who adopted movie m in the "naturally occurred" scenario in our data set during this time period. We then use ΔT_{seed_n} to represent the total number of individuals who adopted movie m from the influence of both "natural" adopters and newly selected seeds from T_1. Therefore, the effects of seeding at T_n can be calculated as $SE_n = (\Delta T_{seed\_n} - \Delta T_n)/\Delta T_n$. Here ΔT_n is used to normalize the seeding effects across movies since each movie has its own intrinsic capability (e.g., popularity) to diffuse. SE_n only measures the additional benefits (adoptions) the seeding strategy (at T_1) brings at time period T_n. Another performance measure we used $TSE_n = \sum_{n=2}^{N}(\Delta T_{seed\_n} - \Delta T_n) / \sum_{n=2}^{N} \Delta T_n$ is to measure the aggregate effects of the applied seeding strategy comparing with the naturally occurred diffusions at the end of the observation period.

In the second setting, we would like to deploy dynamic seeding strategies that not only initiate customers at the start of product release but also other stages/time periods of product diffusion according to our empirical analyses. For instance, as Figure 2 shows that homopihly driven diffusion is increasing at the later stage of "Tinker Tailor Solider Spy" diffusion (from day 105 to 135). For that time period, we may select seeds more based on their similarity with customers who already watched that movie in terms of customer characteristics. The SE_n then measures the effects of such seeding before time period T_n altogether. We also calculate the average TSE_n for all movies to evaluate the overall effectiveness for various seeding strategies.

Conclusion and Intended Contributions

Our study empirically analyzed the moderating impacts of four major types of moderators of peer influence using a large-scale real-world data set. We try to leverage such empirical insights to develop a model that can incorporate the complexities (various moderators) and dynamics (changing peer influence) of product diffusion, as well as realistic data-driven simulations based on it. Our research is among the first to systematically investigate the impacts of the moderators of peer influence and help us better understand how product diffusion happens with different moderator settings. Our empirical model and data-driven simulations help to design and evaluate effective seeding strategies based on specific empirical scenarios. At last, our approach that combines HDPSM as data preprocessing with the social interaction empirical model can also be applied to model and analyze social contagion in other application domains.

References

Abrahamson, E., and Rosenkopf, L. 1997. "Social Network Effects on the Extent of Innovation Diffusion: A Computer Simulation," Organization Science (8:3), pp 289-309.

Adomavicius, G., and Tuzhilin, A. 2005. "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," Knowledge and Data Engineering, IEEE Transactions on (17:6), pp 734-749.

Akbaş, M. İ., Avula, R. N., Bassiouni, M. A., and Turgut, D. Year. "Social network generation and friend ranking based on mobile phone data," 2013 IEEE International Conference on Communications (ICC), IEEE2013, pp. 1444-1448.

Alessio, B., Ming, L., Lonxiang, G., and Wen, S. 2012. "Effects of Social Characters in Viral Propagation Seeding Strategies in Online Social Networks," 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications) 25-27 June 2012, pp 632 - 639.

Aral, S. 2010. "Commentary—Identifying Social Influence: A Comment on Opinion Leadership and Social Contagion in New Product Diffusion," Marketing Science (30:2) 2011/03/01, pp 217-223.

Aral, S. 2011a. "Commentary-identifying social influence: A comment on opinion leadership and social contagion in new product diffusion," Marketing Science (30:2), pp 217-223.

Aral, S. 2011b. "Identifying social influence: A comment on opinion leadership and social contagion in new product diffusion," Marketing Science (30:2), pp 217-223.

Aral, S., Muchnik, L., and Sundararajan, A. 2009. "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks," Proceedings of the National Academy of Sciences (106:51) December 22, 2009, pp 21544-21549.

Aral, S., MUCHNIK, L., and SUNDARARAJAN, A. 2013. "Engineering social contagions: Optimal network seeding in the presence of homophily," Network Science (1:02), pp 125-153.

Aral, S., and Walker, D. 2011. "Creating Social Contagion Through Viral Product Design: A Randomized Trial of Peer Influence in Networks," Management Science (57:9), pp 1623-1639.

Aral, S., and Walker, D. 2012. "Identifying Influential and Susceptible Members of Social Networks," Science (337:6092) July 20, 2012, pp 337-341.

Aral, S., and Walker, D. 2014. "Tie Strength, Embeddedness, and Social Influence: A Large-Scale Networked Experiment," Management Science (60:6), pp 1352-1370.

Arastoo, F., Amir, A., and Ali, J. 2016. "Competitive Diffusion in Social Networks: Quality or Seeding?," (PP:99) 12 April 2016, p 1.

Berger, J., and Milkman, K. L. 2012. "What Makes Online Content Viral?," Journal of Marketing Research (49:2), pp 192-205.

Brock, W. A., and Durlauf, S. N. 2001. "Discrete choice with social interactions," The Review of Economic Studies (68:2), pp 235-260.

Centola, D. 2011. "An experimental study of homophily in the adoption of health behavior," Science (334:6060), pp 1269-1272.

Coleman, J., Katz, E., and Menzel, H. 1957. "THE DIFFUSION OF AN INNOVATION AMONG PHYSICIANS," Sociometry (20:4), pp 253-270.

Cui, K., Zheng, X., and Ieee 2014. "Computational Experiments Based on Competitive Influence Diffusion Model," 2014 Ieee International Conference on Service Operations and Logistics, and Informatics (Soli)) 2014, pp 329-333.

Damon Centola, and Michael Macy 2007. "Complex Contagions and the Weakness of Long Ties," American Journal of Sociology (113:3), pp 702-734.

de Matos, M. G., Ferreira, P., and Krackhardt, D. 2014. "PEER INFLUENCE IN THE DIFFUSION OF IPHONE 3G OVER A LARGE SOCIAL NETWORK," Mis Quarterly (38:4) Dec, pp 1103-+.

Domingos, P., and Richardson, M. Year. "Mining the network value of customers," Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, ACM2001, pp. 57-66.

Eagle, N., Pentland, A. S., and Lazer, D. 2009. "Inferring friendship network structure by using mobile phone data," Proceedings of the national academy of sciences (106:36), pp 15274-15278.

Fang, X., Hu, P. J.-H., Li, Z., and Tsai, W. 2013. "Predicting adoption probabilities in social networks," Information Systems Research (24:1), pp 128-145.

Ghose, A., and Han, S. P. 2011. "An empirical analysis of user content generation and usage behavior on the mobile Internet," Management Science (57:9), pp 1671-1691.

Goldenberg, J., Libai, B., and Muller, E. 2001. "Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth," Marketing Letters (12:3), pp 211-223.

Heath, C., C. Bell, and E.Sternberg 2001. "Emotional selection in memes: The case of urban legends," J. Personality Soc. Psych. (81:6), p 10281041.

Ho, D. E., Imai, K., King, G., and Stuart, E. A. 2007. "Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference," Political analysis (15:3), pp 199-236.
Ibarra, H., and Andrews, S. B. 1993. "Power, social influence, and sense making: Effects of network centrality and proximity on employee perceptions," Administrative science quarterly), pp 277-303.
Jankowski, J., Michalski, R., and Kazienko, P. 2013. "Compensatory Seeding in Networks with Varying Avaliability of Nodes," 2013 Ieee/Acm International Conference on Advances in Social Networks Analysis and Mining (Asonam)) 2013, pp 1242-1249.
Katz, E., and Lazarsfeld, P. F. 1955. Personal Influence: The Part Played by People in the Flow of Mass Communications).
Kempe, D., Kleinberg, J., #201, and Tardos, v. 2003. "Maximizing the spread of influence through a social network," in Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM: Washington, D.C., pp. 137-146.
Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., and Makse, H. A. 2010. "Identification of influential spreaders in complex networks," Nat Phys (6:11) 11//print, pp 888-893.
Koch, O. F., and Benlian, A. 2015. "Promotional Tactics for Online Viral Marketing Campaigns: How Scarcity and Personalization Affect Seed Stage Referrals," Journal of Interactive Marketing (32) Nov, pp 37-52.
Kvasnicka, M. 2014. "Search for the optimal strategy to spread a viral video: An agent-based model optimized with genetic algorithms," Mathematical Methods in Economics (Mme 2014)) 2014, pp 548-553.
Leenders, R. T. A. 2002. "Modeling social influence through network autocorrelation: constructing the weight matrix," Social Networks (24:1), pp 21-47.
Levy, D. A. 1992. "The liberating effects of interpersonal influence: An empirical investigation of disinhibitory contagion," The Journal of social psychology (132:4), pp 469-473.
Levy, D. A., and Nail, P. R. 1993. "Contagion: a theoretical and empirical review and reconceptualization," Genetic, social, and general psychology monographs).
Manski, C. F. 1993. "Identification of endogenous social effects: The reflection problem," The review of economic studies (60:3), pp 531-542.
Oliver Hinz, Bernd Skiera, Christian Barrot, and Becker, J. U. 2011. "Seeding Strategies for Viral Marketing: An Empirical Comparison," Journal of Marketing (75:6), pp 55-71.
Samik, D., Anirban, M., and Nisheeth, S. 2010. "Viral Marketing for Multiple Products," 2010 IEEE International Conference on Data Mining) 13-17 Dec. 2010, pp 118 - 127.
Tucker, C. 2008. "Identifying formal and informal influence in technology adoption with network externalities," Management Science (54:12), pp 2024-2038.
Valente, T. W. 1995. Network Models of the Diffusion of Innovations).
Valente, T. W. 1996. "Social network thresholds in the diffusion of innovations," Social Networks (18:1), pp 69-89.
Van Den Bulte, C., and Joshi, Y. V. 2007a. "New product diffusion with influential and imitators," Marketing Science (26:3), pp 400-421.
Van den Bulte, C., and Joshi, Y. V. 2007b. "New Product Diffusion with Influentials and Imitators," Marketing Science (26:3) 2007/05/01, pp 400-421.
Wang, A., Zhang, M., and Hann, I.-H. 2015. "Socially Nudged: A Quasi-Experimental Study of Friends' Social Influence in Online Product Ratings," Information Systems Research (Forthcoming).
Watts, D. J., and Dodds, P. S. 2007. "Inf luentials, networks, and public opinion formation," Journal of Consumer Research (34:4), pp 441-458.
Wejnert, B. 2002. "Integrating models of diffusion of innovations: A conceptual framework," Annual review of sociology), pp 297-326.
Wellman, B. 1997. "An electronic group is virtually a social network," Culture of the Internet (4), pp 179-205.
Zignani, M., Quadri, C., Gaito, S., and Rossi, G. P. 2015. "Calling, texting, and moving: multidimensional interactions of mobile phone users," Computational Social Networks (2:1), p 1.